Electronics and Communications in Japan, Part 1, Vol. 79, No. 1, 1996
Translated from Denshi Joho Tsushin Gakkai Ronbunshi, Vol. 77-B-II, No. 12, December 1994, pp. 767-779

p. 82-90 = (15)

433a Electronics and Communications in Japan
Part I: Communications
79(1996) January, No. 1, New York, US

Phase-Tracking Viterbi Demodulator

Mutsumu Serizawa, Shigeru Asakawa,* and Junzo Murakami,** *Members*

Research & Development Center, Toshiba Corporation, Kawasaki, Japan 210

SUMMARY

The maximum-likelihood sequence estimation reception system, where the transmitted signal sequence is decided by the maximum likelihood using Viterbi algorithm, from the modulated signal, such as the phase modulated convolution-coded signal, TCM, and CPM, is considered to be very interesting because of its excellent signal-to-noise ratio (SNR) versus error rate performance. Such a system assumes the coherent detection of the received signal in order to realize excellent performance. The reception system, which is expected to follow the carrier with a large offset and a low SNR, has been difficult to realize due to the difficulty of satisfying the requirement. This paper proposes the following system. The phase-locked loop is provided corresponding to each survival path in the Viterbi algorithm used in the estimation of the maximum-likelihood sequence estimation. The phase-locked loop is selected in parallel to the selection of the survived path. In other words, the phase synchronization and the (maximum-likelihood) sequence estimation are simultaneously achieved. The system is discussed in detail, and it is shown that a stable coherent detection can be realized for a very low SNR.

Key words: Satellite communication; mobile communication; maximum-likelihood sequence estimation; convolutional code; phase-shift keying.

*Presently with Kansai New Technology Research Laboratory.

*Presently with Toshiba Research Consulting Co.

1. Introduction

The maximum-likelihood sequence estimation reception system is considered in which the phase modulation of the convolution-coded signal [1, 2], the continuous phase modulation (CPM) [3-7], the trellis-coded modulation (TCM) [8, 9], and the modulation with intersymbol interference [10, 11] are used; and the transmitted signal sequence is estimated by the maximum-likelihood decision using the Viterbi algorithm, from such a modulated signal in the wide sense, containing coding. Such a system is considered very interesting, because of its excellent SNR vs. error-rate performance. The system exhibits an excellent spectrum efficiency and the error rate performance. It is an especially useful modulation system in the satellite communication and the mobile communication, in which severe requirements are imposed in the band and power limitations. The maximum-likelihood sequence estimation technique is important in such a system.

The maximum-likelihood sequence estimation using Viterbi algorithm is usually applied to the baseband signal, which is obtained by the coherent detection of the received signal [1, 2, 10]. In the satellite communication, the demodulator must work for the low SNR due to the severe limitation of the transmitter power. The demodulator must work for a still lower SNR in applications such as deep-space exploration where the channel coding with a high error-correcting ability, such as the convolutional code with a long constraint length or the concatenation code, is used. In the case of the cellular mobile communication system, it is required that the receiver should work correctly for the severe condition (such as low CIR

ISSN8756-6621/96/0001-0082

low CIR and SNR), in order to reduce the number of cell iterations and to utilize the spectrum effectively. One of the major reasons in the deterioration of the receiver performance in the low SNR condition is the generation of the unreceivable state due to the cycle slip of the regenerated carrier, in addition to the increase of the error rate due to the phase jitter of the regenerated carrier. The absolute-coded phase modulation, without the differential coding, exhibits a better SNR vs. error rate performance, compared to the case using the differential coding; it is widely used in the satellite communication system where the transmitter power is severely limited.

In the system without differential coding, the known signal is inserted at regular intervals in the transmitted sequence in order to indicate the absolute phase of the transmitted signal. Based on that signal, the phase uncertainty is eliminated and the absolute phase is established. When a desynchronization occurs, the demodulator cannot work correctly until the synchronization is re-established at the same absolute phase before the desynchronization (error rate = 0.5). Even if the synchronization is re-established after the desynchronization, the demodulator does not work correctly until it receives the next known signal if there is produced a cycle slip in which the synchronization is established at an absolute phase which is different from the phase before the desynchronization. Thus, after receiving the known signal, the phase synchronization must be maintained in the coherent detector of the receiver until the next known signal is received while suppressing the generation of the cycle slip. In other words, the probability for the cycle slip generation should be suppressed to a very low value, especially when the receiver must operate for a low SNR.

In the satellite communication or the mobile communication, there exists a large frequency offset or the carrier phase variation, due to the inaccuracy of the synthesizer frequency, the fading or the Doppler shift due to the motion of the satellite. Consequently, the coherent detector to be used in those systems must follow the large and rapid variation of the carrier phase.

A typical method to synchronize the regenerated carrier to the received signal follows. The coherent detection using the regenerated carrier is applied to the received signal. The phase error is extracted by eliminating the modulation component from the obtained baseband signal. Using the signal obtained by eliminating the noise component from the error signal by the loop filter, the phase of the regenerated carrier is controlled. The noise immunity and the follow-up performance to the phase variation in such a system are determined by two basic functions: one to eliminate the modulation component from the received signal (the phase error detection function), the other to eliminate the bandwidth of the loop filter or the tank circuit.

In general, it is considered that the loop filter of the coherent detector should be designed to be wide, in order to realize a wide pull-in frequency range and to follow up more easily to the phase variation (such as the fading and Doppler shift). On the other hand, the bandwidth of the loop filter in the coherent detector should be sufficiently narrow in order that the detector can operate for a low SNR. In other words, it is difficult to satisfy these two requirements, i.e., good noise immunity and good follow-up to phase variation, by manipulating only the parameter of the loop filter.

As to the other function, i.e., the elimination of the modulation component, there is no problem if the modulation signal component is known. However, it is unknown in the coherent detector that tries to detect the transmitted signal. Consequently, in the conventional system, the modulation component is estimated by some means and the phase error signal is derived based on that estimation. Then the performance of the modulation component estimation is the decisive factor in determining the performance of the phase detector. In such a conventional system, the error rate of the transmitted signal estimation in the output of the coherent detector is very high for the low SNR. This prevents satisfactory detection of the phase error and causes deterioration of the coherent detection performance.

Other than the foregoing system, where the coherent detection and the maximum-likelihood estimation are separately provided, there are efforts to improve the receiver performance by integrating those functions [12-15]. Such a system, however, has the following problems. The sequence estimator with a large delay is included in the coherent detector which deteriorates the loop response and prevents the realization of a satisfactory follow-up to the phase variation [12, 13]. The carrier phase is handled as the state in the Viterbi algorithm and the receiver performance is greatly deteriorated when there is a large frequency offset [14, 15].

From such a viewpoint, this paper proposes the following system where the phase synchronization and the (maximum-likelihood) sequence estimation are simultaneously executed. The phase-correction loop is provided corresponding to each of the survived paths in the Viterbi algorithm to be used in the maximum-likelihood

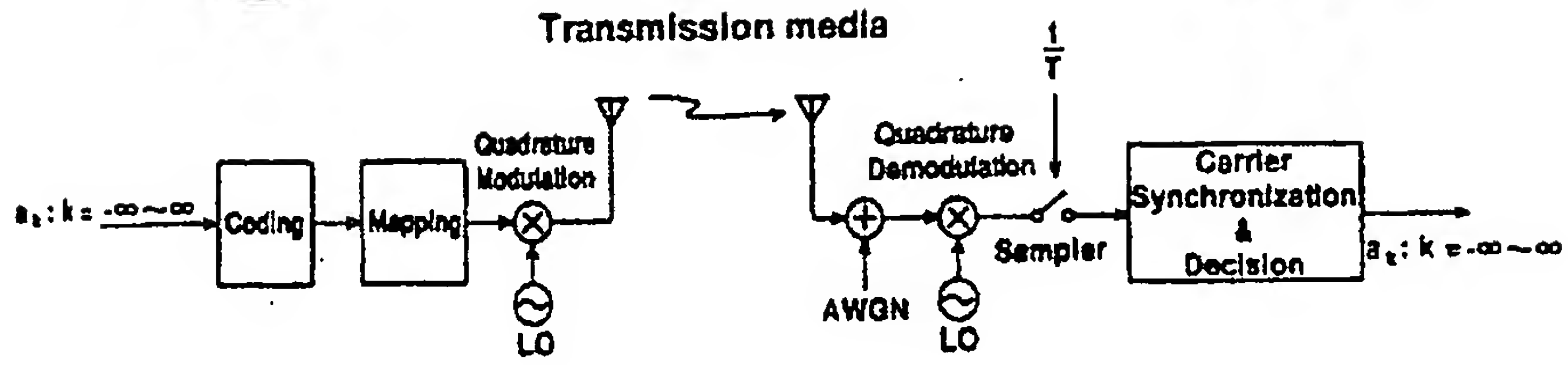

Fig. 1. Transmission system model.

sequence estimation. Then the phase-correction loop is selected simultaneously with the survived path. In this system, the modulation component can be estimated with a high probability even for a low SN condition, and it is expected that the synchronization performance will be greatly improved for the low SNR. A special advantage is that a large delay is not included in the phase-correction loop; the follow-up performance is satisfactory even if there is a large phase variation.

The proposed system is discussed in detail in the following. It is shown by simulation that an excellent demodulation performance is obtained when the method is applied to the phase-modulated convolutionally coded signal, even if the SNR is low and there is a large frequency offset. Section 2 presents the channel model and the problems in the conventional coherent detector. Section 3 outlines the method to integrate the maximum-likelihood sequence estimation and the phase detector. Section 4 discusses the application of the Viterbi algorithm. Section 5 presents the result of examination by simulation.

2. Transmission System Model and Conventional Coherent Detector

2.1. Transmission system model

The transmitter/receiver system is modeled as in Fig. 1. The transmission data sequence $\{a_i: i = -\infty \sim \infty$, a_i is rM bit data$\}$ is encoded (with coding ratio r) and modulated (M bit/symbol). It is affected by certain channel characteristics as well as by the additive Gaussian noise (AGWN). The quadrature demodulation is applied using the asynchronous carrier. The signal is sampled by the symbol rate ($1/T$), and the sampled sequence $\{z_i: i = -\infty \sim \infty\}$ is obtained. It is assumed that the sampling clock is synchronized in phase to the transmitter clock. The task of the proposed system is to estimate the transmitted data sequence $\{a_i: i = -\infty \sim \infty\}$ from the sampled value sequence $\{z_i: i = -\infty \sim \infty\}$.

2.2. Received sampled sequence

It is assumed that the input/output characteristics (constraint length ν) of the modulator and the encoder, as well as the impulse response lengths of the channel and the transmitter/receiver filters, are finite. The sampled sequence is given as follows:

$$z_i = f_i(a_{i-\nu+1}, \cdots, a_i) * \exp(-j\theta(iT)) + n' \quad (1)$$

where $f_i(a_{i-\nu-1}, \cdots, a_i)$ is the overall transmission characteristics, determined by the coding method, modulation method, and the impulse responses of the transmitter/receiver filters. $\theta(iT)$ is the phase difference between the unmodulated carrier and the local oscillator output of the receiver. n' is the complex Gaussian noise affected by the receiver filter. When the transmitted signal is the phase-modulated convolutional code, $f_i(a_{i-\nu-1}, \cdots, a_i)$ takes one of the following values.

$$\exp(-jn\pi/2^{M-1}) : n = 0, 1, \cdots, 2^M - 1$$

The conventional modulation/demodulation system works as follows. Using the coherent detector, the above phase component $\exp(-j\,\theta(iT))$ is compensated and modulation-eliminated. Based on the obtained demodulated baseband signal

$$\begin{aligned} S_r &= z_i \exp\{j[\theta(iT) + \delta\phi]\} \\ &= f_i(a_{i-\nu+1}, \cdots, a_i)\exp(j\delta\phi) \\ &\quad + n'\exp\{j[\theta(iT) + \delta\phi]\} \end{aligned} \quad (2)$$

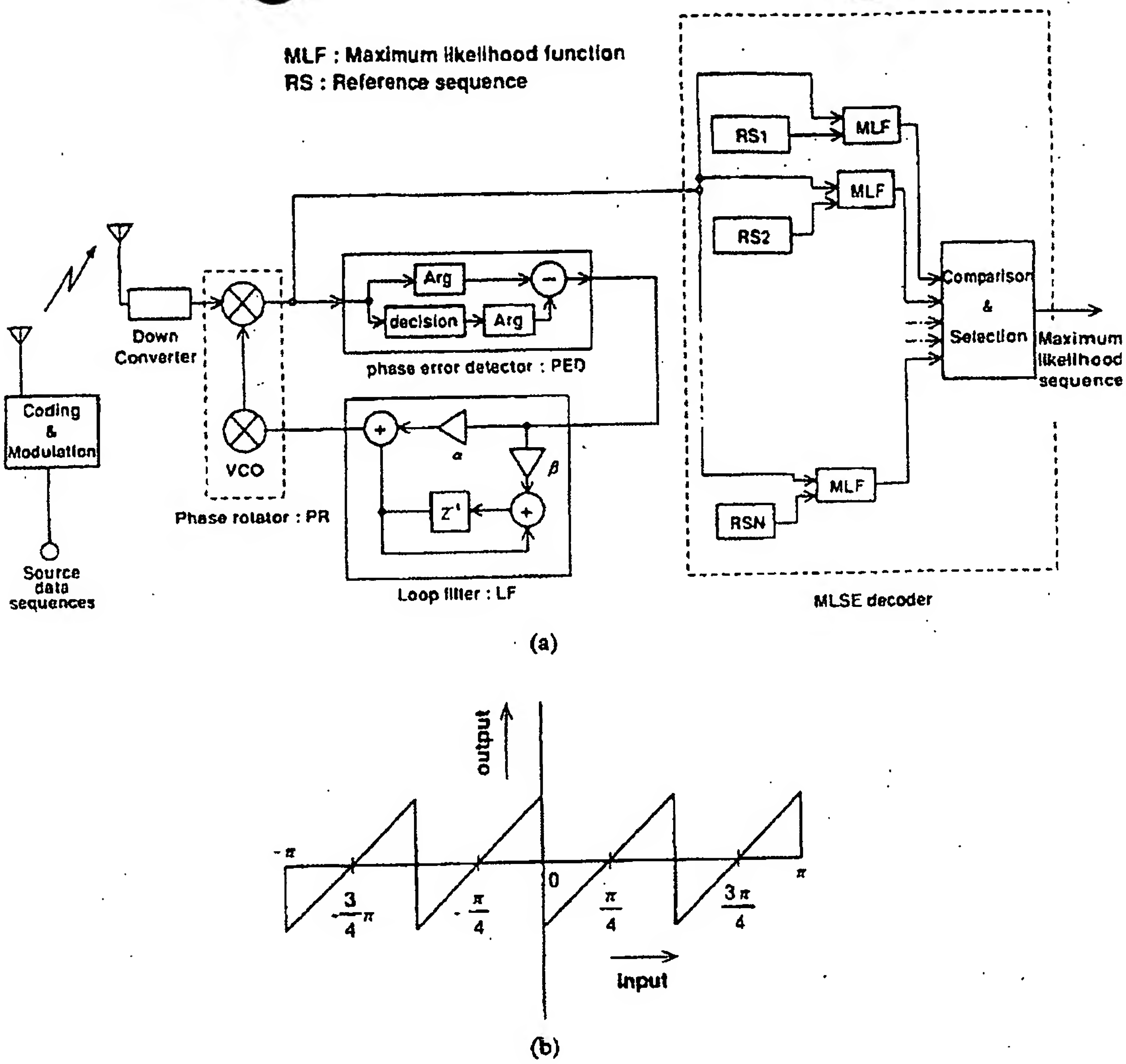

Fig. 2. (a) Conventional QPSK coherent demodulator and decoder; and (b) phase error detector characteristic of conventional QPSK coherent demodulator.

the transmitted data sequence $\{a_i;\ i = -\infty \sim \infty\}$ is estimated using the maximum-likelihood sequence estimation. In the forementioned, $\delta\phi$ is the phase error contained in the output of the regenerated carrier oscillator. When the SNR is sufficiently high in general, there holds

$$\mathrm{Abs}(\delta\phi + n^*) \ll \pi/2^{M-1} \tag{3}$$

in the steady state of the coherent detector. In the foregoing, Abs(x) is the absolute value of x. n^* is the noise phase component due to the effect of the noise $n'\exp\{j[\theta(iT) + \delta\phi]\}$.

2.3. Conventional coherent detection and problems

Figure 2(a) shows conceptually the conventional coherent detection system. As is shown in the figure, the conventional coherent detector extracts the phase error of the regenerated carrier from the demodulated baseband signal and the phase of the regenerated carrier is controlled by the phase error signal. Usually, the phase error of the regenerated carrier is extracted by deciding the demodulated baseband signal as one of $\exp(-jn\pi/2^{M-1})$: $n = 0, 1, 2, \cdots, 2^M - 1$ and by determining the phase difference between the decided value and the

demodulated baseband signal. In practice, the transmitted signal is coded and the signal cannot always take any value of $\exp(-jn\pi/2^{M-1})$: $n = 0, 1, 2, \cdots, 2^{m} - 1$.

The phase error is thus obtained by passing the received baseband signal Sr through the periodic sawtooth phase response with zero values at $n\pi/2^{M-1}$: $n = 0, 1, 2, \cdots, 2^{M} - 1$ (as is shown in the phase error detector of Fig. 2(b) for $M = 2$). In the past, the 2^{M}-multiplier is known as a method for realizing such a response. The result obtained by such a response is a little different, but is essentially the same.

The foregoing method has no problem in the operation when the SNR is high enough to guarantee Eq. (3). In other words, when the SNR is sufficiently high, there is little decision error in the value decided by the coherent detector. The phase difference between the decided value and the received baseband signal then is the sum $\delta\phi + n^{*}$ of the phase error component and the noise. By deleting the noise component from the sum signal using a filter, the phase error component can be extracted with a sufficient accuracy.

When the SNR is low, on the other band, the decided value may contain various errors. When the four-phase PSK operates near E_b/N_0, the error rate is approximately 10 percent. In this case, the detected phase error contains the detection error due to the decision error by a ratio of 10 percent. In the case of a decision error, a value entirely different from $\delta\phi + n^{*}$ is inputted from the phase error detector.

Consider, for example, the case where a decision error is produced in the four-phase PSK when the error signal exceeds only slightly the decision threshold. If the phase error detection response of Fig. 2(b) is used in such a case, the magnitude and the sign of the detected phase error signal greatly differs, depending upon whether the baseband signal exceeds or does not exceed the threshold. In other words, instead of extracting the phase error $\delta\phi + n^{*}$, $\delta\phi + n^{*} \pm \pi/2$: $\mathrm{Abs}(\delta\phi + n^{*} \pm \pi/2) \leq \pi/4$ may be outputted. This detection error is a nonlinear "noise," which produces a large deviation from the true phase error. Thus, in the widely used conventional coherent detector, this nonlinear noise drastically deteriorates the phase synchronization performance when there are produced a large number of decision errors in the phase error detector. This is the factor that prevents the satisfactory operation of the system in low SN condition.

When the convolution code of constraint length 7 and rate 1/2 is used in the four-phase PSK, even if there is an error of some 10 percent before the decoding, the error rate is reduced to approximately 0.1 percent after the decoding. In the communication system that operates in practice around this error level, the coherent detector must cope with the error rate around 10 percent. In other words, there must be a coherent detector that realizes a satisfactory operation in the adverse condition as described here. This implies that a more severe condition is imposed on the coherent detector as the error-correcting ability is more improved.

2.4. Coherent detector for low SNR

Using the foregoing reasoning, it can be seen that one of the factors that prevents the satisfactory operation of the coherent detector for the low SNR is the decision error in the phase error extraction in the coherent detector. One way to remedy the performance deterioration due to the decision error may be to utilize positively the convolution code multiplied with the transmitted signal and the maximum-likelihood sequence estimation to reduce the decision error.

In general, it is inevitable that a considerable delay is produced in the system to decode or demodulate the convolutional code or the intersymbol interference. When the maximum-likelihood sequence for the convolutional code is estimated by the Viterbi algorithm, a decoding delay of several times longer length compared to the constraint length is considered as necessary, especially for the low SN condition.

If such a maximum-likelihood sequence estimator is directly applied to the decision unit of the coherent detector, a very large delay is included in the phase-locked loop. A large delay in the loop affects the stability of the loop and deteriorates the synchronization performance. In other words, a large improvement of the performance cannot be expected if the maximum-likelihood sequence estimator is directly included in the coherent detector. In the following sections, a method is proposed and investigated in which the maximum-likelihood sequence estimation with the low error rate is applied to the decision-making in the coherent detector without deteriorating the synchronization performance.

3. Maximum-Likelihood Sequence Estimator with Phase-Tracking Function

In order to help understand the maximum-likelihood sequence estimation method using the Viterbi algorithm discussed in the next section, assumptions are

presented in different stages, together with some accompanying discussions. In the following, section 3.1 discusses the sequence estimation assuming a complete phase synchronization. Section 3.2 describes the phase (correction) synchronizer assuming that the transmitted sequence is known. Section 3.3 discusses the sequence selection, after applying the phase correction to each candidate sequence that has the possibility of having been transmitted.

The first step is to consider the estimation of the transmitted sequence a_i: $i = 1, \cdots, p$ from the sampled sequence z_i: $i = 1, \cdots, p$. The maximum-likelihood estimation is the following procedure. When a signal is received, the transmitted data sequence with the highest probability (likelihood) is sought, among all data sequences (2^p in this case) that have the possibility of having been transmitted. The data sequence with the highest probability is decided as the transmitted sequence. In other words, in the maximum-likelihood decision, the likelihood is calculated for each of the data sequences that has the possibility of being transmitted and the transmitted data sequence is selected as the sequence with the highest probability.

The sampled sequence z_i: $i = 1, \cdots, p$ is the output from the quadrature detector using the local oscillator that is not phase synchronized. It contains the phase difference $\theta(iT)$ between the carrier and the local oscillator output of the receiver and cannot directly be used in the estimation of the transmitted sequence a_i: $i = 1, \cdots, p$. To estimate the transmitted sequence a_i: $i = 1, \cdots, p$, it must be the case that either the phase difference $\theta(iT)$ is known or is estimated by some means.

3.1. Sequence estimation assuming a complete synchronization

As the first step, consider the case where a complete synchronization is established between the phase of the carrier and the phase of the local oscillator of the receiver.

$$\theta(iT) = 0 \tag{4}$$

Then the likelihood of each of the $h = 2^p$ data sequences that may have been transmitted can be determined as follows. Assuming that each sequence is transmitted, the distance between the actual received signal and the signal point (reference signal) that will be obtained if the ideal signal is received, is calculated.

Then the likelihood can be determined uniquely based on the sum of the square distances:

$$L_h = \sum_{i=1}^{p} |z_i - r_i^h|^2 \tag{5}$$

The likelihood is high if the foregoing square-sum distance is smaller. Here, r_i^h is each of the $h = 2^p$ signal sequences that have the possibility of having been transmitted. There are obtained correspondingly 2^p likelihoods. Among those 2^p likelihoods, the maximum is determined and the sequence corresponding to the maximum is outputted as the sequence with the maximum likelihood.

The calculation of the likelihood shown in Eq. (5) based on the distance and the sequence estimation based on the likelihood correspond to the maximum likelihood sequence estimation in the state where the conventional coherent detector is operating correctly. As was discussed, the synchronization performance of the conventional system in Fig. 2(a) deteriorates when there is a decision error in the decision unit in the phase error detector, which prevents the calculation of the likelihood shown in Eq. (5) as well as the sequence estimation based on the result.

3.2. Phase synchronizer assuming known transmitted sequence

As a next step, consider the case in which the complete phase synchronization is not established between the local oscillator of the receiver and the carrier. Assume that the transmitted signal is known in the receiver by some means. In this case, the known transmitted signal sequence can be used instead of the output of the decision unit in Fig. 2(a). This corresponds to the case in which no error is included in the decision output.

Then, corresponding to the transmitted data sequence, the sequence ($\theta(iT)$: $i = 1, \cdots, p$) of the phase difference between the local oscillator output and the carrier can be estimated. The estimated value for $\theta(iT)$ is the output of the loop filter when the loop of the coherent detector of Fig. 2(a) is cut at the output of the loop filter. Estimation for our purposes means that the effect of the noise is considered. The phase error sequence can be estimated if the SNR is infinite. When the SNR is infinity, the phase sequence is represented as follows:

$$\theta(iT) = \mathrm{Arg}(z_i) - \mathrm{Arg}(r_i) \tag{6}$$

The synchronizer operates in order to compensate the foregoing phase difference. This corresponds to the case in which the loop of the coherent detector is closed. Then the output phase of VCO tends to compensate $\theta(iT)$. In practice, however, the transmitted data sequence is unknown, which is the original problem. This point is considered in the next section.

3.3. Sequence selection after phase correction of each sequence candidate with possibility of transmission

Assume that a sampled sequence z_i: $i = 1, \cdots, p$ is obtained. There 2^p data sequences that have the possibility of having been transmitted. For the given sampled sequence, an arbitrary one of the 2^p data sequences that have the possibility of having been transmitted is selected and the phase synchronizer is operated. This is the same as operating the coherent detector, using an assumed sequence r_i^* of 2^p sequences instead of the decision output (in PED) of Fig. 2(a). In this case, the coherent detector operates to compensate the phase error sequence $(\theta^h(iT)$: $i = 1, \cdots, p)$

$$\theta^h(iT) = \mathrm{Arg}(z_i) - \mathrm{Arg}(r_i^h) \tag{7}$$

based on the assumed sequence. The output of this coherent detector is the phase corrected input signal sampled sequence z_i^h

$$z_i^h = z_i * \exp(+j\phi_i^h) \tag{8}$$

In the foregoing, ϕ_i^h is the phase generated inside of the coherent detector so that the coherent detector compensates the phase error sequence $\theta^h(iT)$. It is obtained by integrating the output of the loop filter in the coherent detector. In other words, ϕ_i^h is the result of deleting the nonstationary component, such as noise, from $\theta^h(iT)$.

Such a phase-corrected input signal sampled sequence can be determined for each of the $h = 2^p$ assumed sequences. In other words, $h = 2^p$ phase-corrected input signal sampled sequences can be determined. One of those $h = 2^p$ assumed sequences is the true transmitted sequence. In other words, one of the 2^p phase-corrected input signal sampled sequences is the input signal sampled sequence, which is the result of coherent detection/ phase correction based on the true transmitted sequence.

The likelihood can be calculated for each phase-corrected input signal sampled sequence. In other words, the likelihood for the h-th sequence is determined from the phase-corrected input signal sampled sequence derived based on the h-th sequence, together with the reference signal for the h-th sequence. The likelihood is calculated based on the square-sum of the distance

$$L_h^* = \sum_{i=1}^{p} |z_i \cdot \exp(j\phi_i^h) - r_i^h|^2 \tag{9}$$

The likelihood is defined as large when L_n^* is small.

If the "assumed transmitted data sequence h" is the same as the actually transmitted data sequence, the phase-corrected input signal sampled sequence z_i^* is the same as the sampled sequence when the phase synchronization is established between the local oscillator of the receiver and the carrier. In that case, the square-sum of the distance in Eq. (9) is the same as that of the case where the transmitted sequence is assumed as known, as in section 3.2.

When the likelihood is calculated based on Eq. (9), for all possible transmitted data sequences, one can estimate for the received signal which of the possible transmission sequences has the highest probability of having been transmitted. At least, if $SNR = \infty$, the maximum likelihood is obtained for the actually transmitted sequence. Also, in the actual case where $SNR \neq \infty$, the likelihood for the data sequence calculated from the phase-corrected sampled sequence based on the correctly assumed transmitted data sequence is the maximum with a high probability compared to the case where other sequences are assumed.

By such a scheme, the estimation of the transmitted data sequence and the estimation of the phase error sequence, i.e., the maximum likelihood estimation and the phase synchronization, can be executed simultaneously. Figure 3 shows the proposed system where the phase synchronization and the sequence estimation are simultaneously executed. The principle is realized by providing 2^p phase-correction loops. In Fig. 3, the phase-error detector (PED) in the individual phase-correction loop differs slightly from the PED shown in Fig. 2(a). The difference is that PED in Fig. 2(a) contains internally the decision unit while that of Fig. 3 does not contain a decision unit and the reference signal (RS) is used instead of the decision output.

In such a system, the phase synchronization is maintained for at least one of a very large number of assumed transmitted data sequences, which is the result

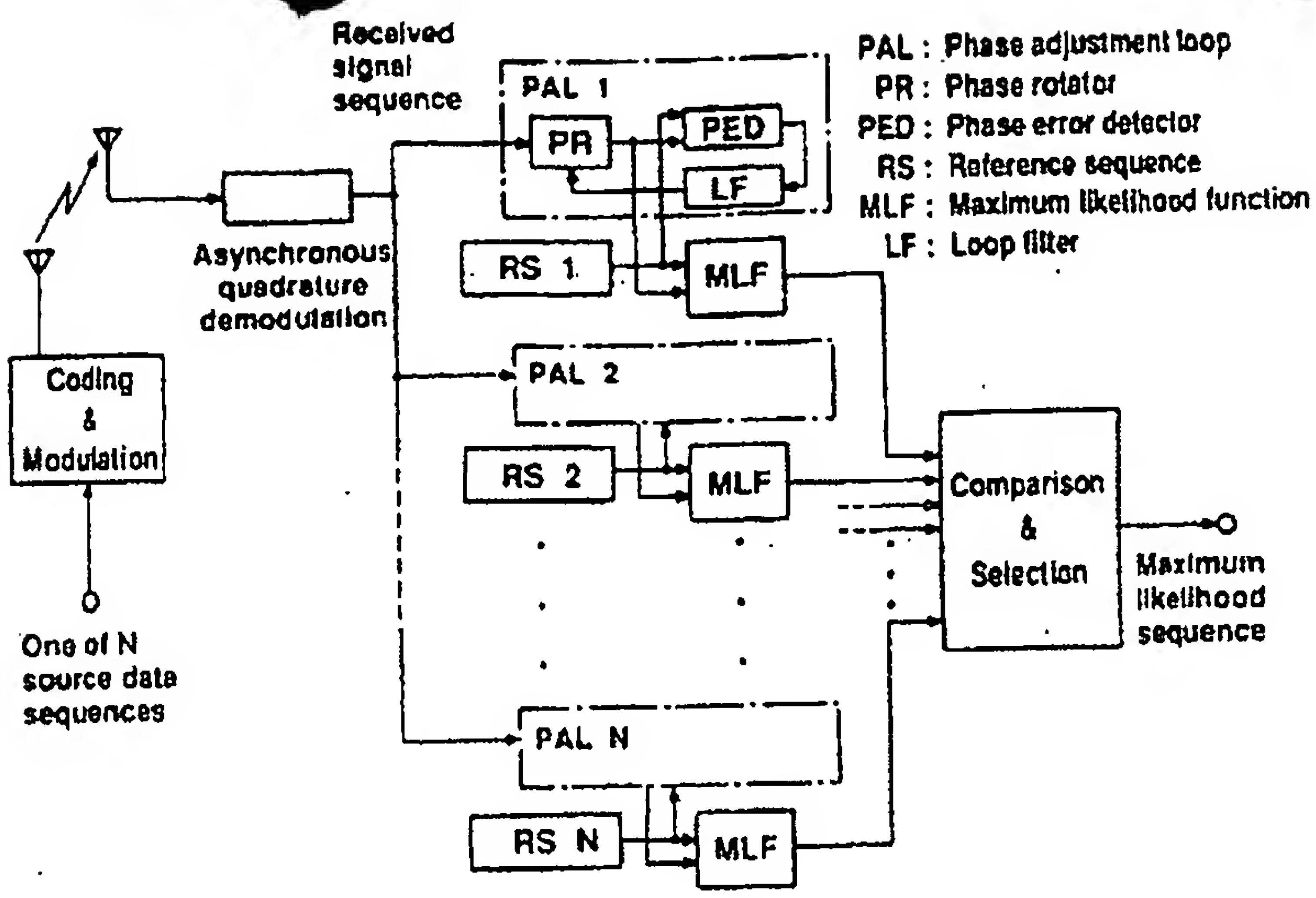

Fig. 3. Transmitted sequence estimation scheme with phase-adjustment loop.

of the phase correction of the actually received signal, assuming the maximum-likelihood transmitted data.

4. Phase-Tracking Viterbi Demodulation Algorithm

If the foregoing series of operations are directly implemented, a tremendous amount of transmission data sequences and the phase-correction loops must be prepared, which will not be realizable. To realize the same operation with less hardware, the Viterbi algorithm is applied. Viterbi algorithm is the algorithm that retains the specified number of survived paths and decides the transmission sequence with the maximum likelihood by gradually updating the path metric for each path.

To combine the phase-tracking function with the conventional Viterbi algorithm, the following two mechanisms are newly added.

① The phase-correction loop function to correspond to each survived path.

② ACS operation is applied based on the calculated branch metric. The path as well as (the internal state of) the phase-correction loop are selected so that the reference signal for the new state and (the internal state of) the phase-correction loop are generated.

Let there be at least 2^{v-1} states of the transmission code trellis (v is the constraint length). The individual phase-correction loop corresponds to each of the survived paths of the state. The phase-correction loop is considered as the secondary loop. Consequently, the internal state of the phase-correction loop at time i can be completely represented by the phase correction ϕ_i^h: $h = 1, 2^{v-1}$ and the frequency offset f_i^h: $h = 1, 2^{v-1}$, which is the first-order derivative of the phase correction. In each phase-correction loop, the following phase rotation is applied to each input sampled value based on the phase correction ϕ_i^h:

$$z_{i+1}^h = z_{i+1} * \exp(+j\phi_i^h) \tag{10}$$

Thus, different phase corrections are applied to the obtained states. Based on the input sampled sequence after the phase correction, the branch metric $Bm_{i,i+1}^{h,k\to m}$ is calculated as

$$Bm_{i,i+1}^{h,k\to m} = |z_{i+1}^h - r_{i,i+1}^{h,k\to m}|^2 \tag{11}$$

In the forementioned, $Bm_{i,i+1}^{h,k\to m}$ is the branch metric corresponding to the branch from state k to state m in the shift from the i-th sample to the $(i+1)$-th sample. $r_{i,i+1}^{h,k\to m}$ is the reference signal corresponding to the transition from state k to state m, in the shift from the i-th sample to the $(i+1)$-th sample. It is calculated in the receiver for the ideal case.

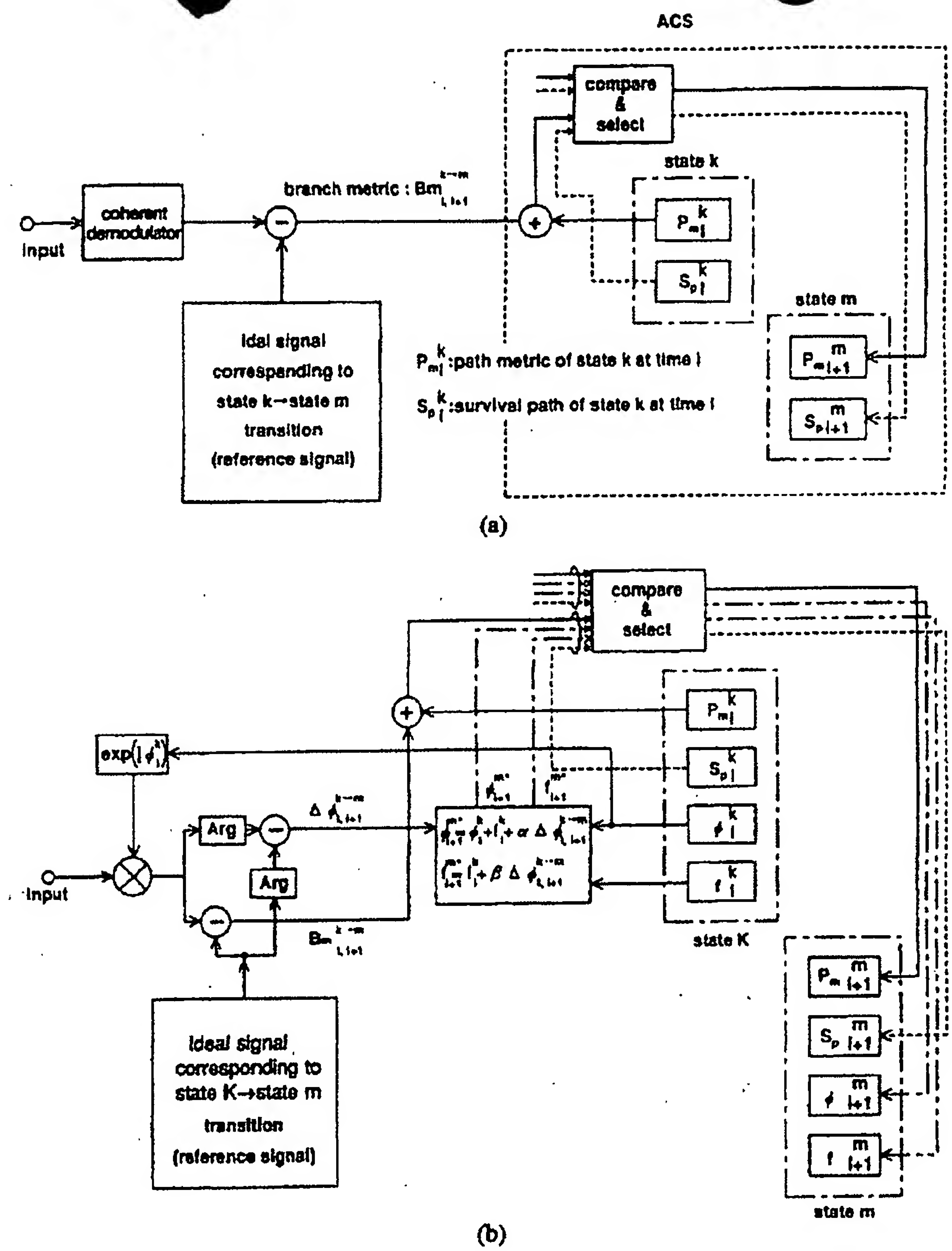

Fig. 4. (a) Basic calculation of conventional Viterbi algorithm, and (b) basic calculation of phase-tracking Viterbi demodulator.

In parallel to the calculation of the branch metric, the phase error data $\Delta\phi_{l+1}^{k\to m}$ is determined as follows:

$$\Delta\phi_{l+1}^{k\to m}=\mathrm{Arg}(z_{l+1}^{k})-\mathrm{Arg}(r_{l+1}^{k\to m}) \qquad (12)$$

where $-\pi<\Delta\phi_{l+1}^{k\to m}\leq\pi$,

Based on the path metric for each state and the branch metric obtained here, the ACS operation is applied to determine the new path. The path metric of each state is updated as well as the phase correction. The phase correction ϕ_{l+1}^{m} is controlled and memorized as follows, based on the phase error $\Delta\phi_{l+1}^{k\to m}$ determined for the transition from state k to state m.

$$\phi_{i+1}^{m} = \phi_i^h + f_i^h + \alpha \Delta \phi_{i+1}^{h \to m} \tag{13a}$$

$$f_{i+1}^{m} = f_i^h + \beta \Delta \phi_{i+1}^{h \to m} \tag{13b}$$

Since the phase-tracking loop is considered as the secondary loop, the frequency data f_i^h is also memorized as the internal state of the loop, in addition to the phase-correction data. α and β correspond to the parameters of the loop filter. They have the same physical significance as the parameters of the loop filter in the conventional coherent detector shown in Fig. 2(a).

Figures 4(a) and (b) show conceptually the conventional Viterbi algorithm as well as the stepwise ACS operation processing of the proposed Viterbi demodulation algorithm with the phase-tracking function. As is seen from the figure, the path is determined by selecting the larger attached path metric. The branch metric required in this process is calculated based on the sampled value z^h, which is phase-corrected by the carrier data in the phase-correction loop corresponding to the survived path h arriving at the start state of that branch metric as well as the reference signal, which is newly calculated using the reference signal corresponding to the survived path arriving at the start state of that branch.

The branch metric is calculated for more than one (two in general, as in the case of the convolutional code) branches arriving at a state. The result is added to the path metric of each start state of those multiple branches. The results are compared, and one of the branches is selected as the survived path. As is described before, the phase correction and the reference signal are determined corresponding to the selected path. Then, using the transmitted data corresponding to the selected branch, the phase error signal is calculated from the sampled value and the phase correction is applied to the start state of the selected branch. The result is used as the new carrier data for the survived path.

By applying the Viterbi algorithm, the data sequences which obviously have smaller possibilities of having been transmitted compared to other sequences, are gradually discarded. At the same time, the synchronization loops corresponding to those sequences are also discarded. By this scheme, the same performance as the phase tracking and sequence estimation system shown in Fig. 3 is obtained.

5. Simulation

The validity of the proposed method is verified by a computer simulation. It is shown in the following that the proposed method where the maximum-likelihood estimator containing the phase-tracking function can realize a stable demodulation in the very low SN condition. The overall error rate performance can be improved, compared to the conventional system outside of the receiver, where the coherent detector and the maximum-likelihood sequence estimator are serially placed and operate independently.

5.1. Simulation condition

The case is considered where the transmitted signal sequence is convolutionally encoded and is transmitted by the absolute phase four-phase PSK. The performance of the proposed method is examined by a computer simulation. The convolutional code of constraint length 5 and rate 1/2 is used. The 2-bit output signal produced for each input bit to the encoder are mapped to the in-phase and quadrature channels of the same timing in the four-phase PSK. As the transmitter/receiver filters, the root-roll-off filter (roll-off factor 30 percent) is assumed, and the Nyquist response is assumed for the overall response of the transmission channel. The SNR is always represented by the carrier-to-noise ratio (CNR) after passing through the receiver filter.

The known signal is inserted once in each 1000 symbols of the modulated signal (1000 bits of the signal before encoding), and the phase uncertainty is assumed to be eliminated using the known signal. Consequently, 1000 symbols at the maximum produce the undecodable burst error when there is produced a cycle slip or the normal phase tracking is made impossible. In the following, the above 1000 symbols plus the known signal is defined as a slot.

In the phase-tracking Viterbi algorithm proposed in this paper, the phase correction corresponding to the survival path is selected. Consequently, the phase uncertainty, which has been a problem in the conventional absolute phase four-phase PSK and in other similar systems, is eliminated. In other words, the known signal to eliminate the phase uncertainty is unnecessary in principle. In the conventional system, which is to be compared to the proposed system, the known signal must be inserted in order to eliminate the phase uncertainty. In addition, the synchronization performance is in general improved by inserting the known signal. As a result, it is assumed that the known signal is inserted into the modulated signal so that the conventional system and the proposed system can be compared under the same condition.

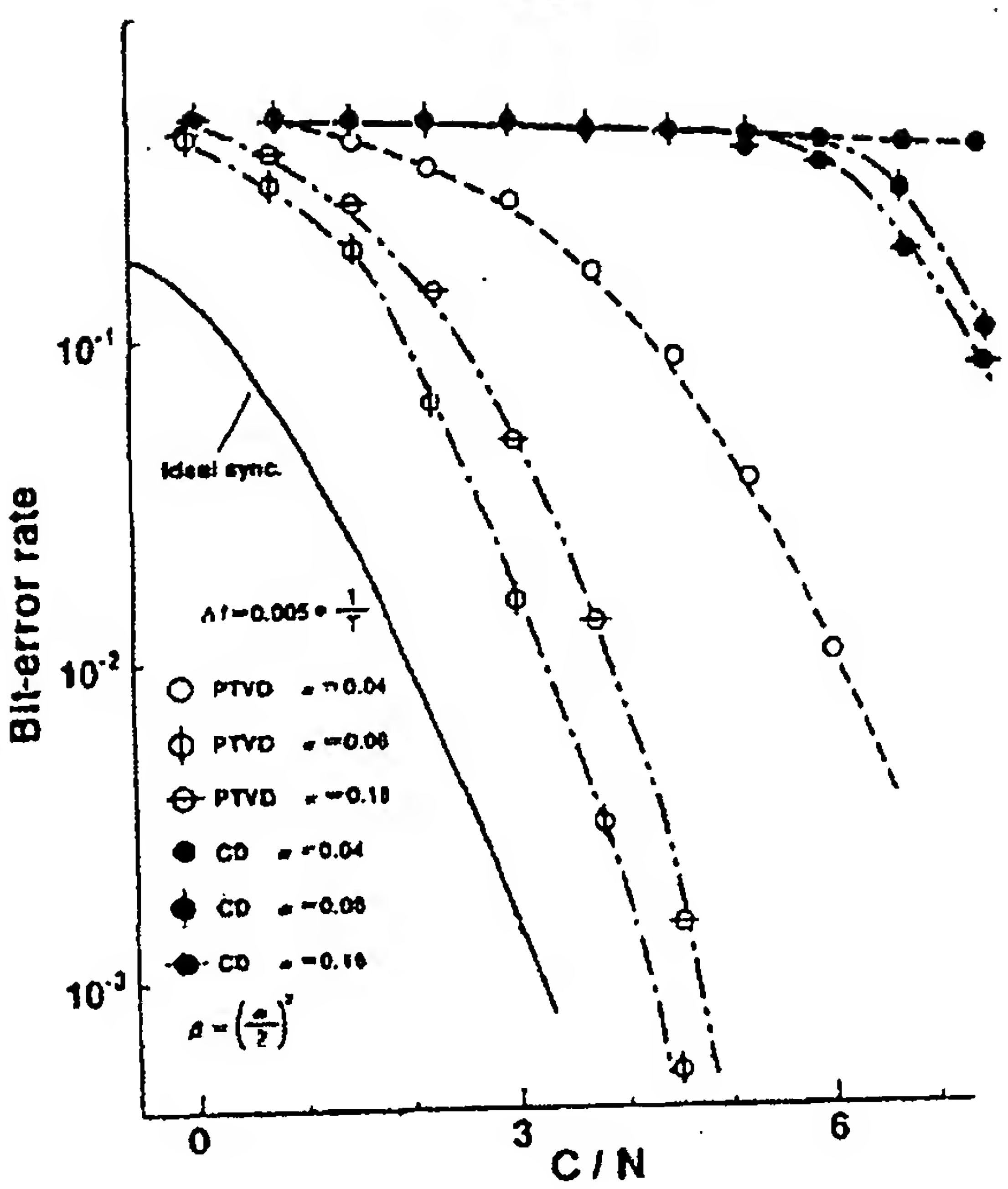

Fig. 5. Bit-error rate for phase-tracking Viterbi demodulator.

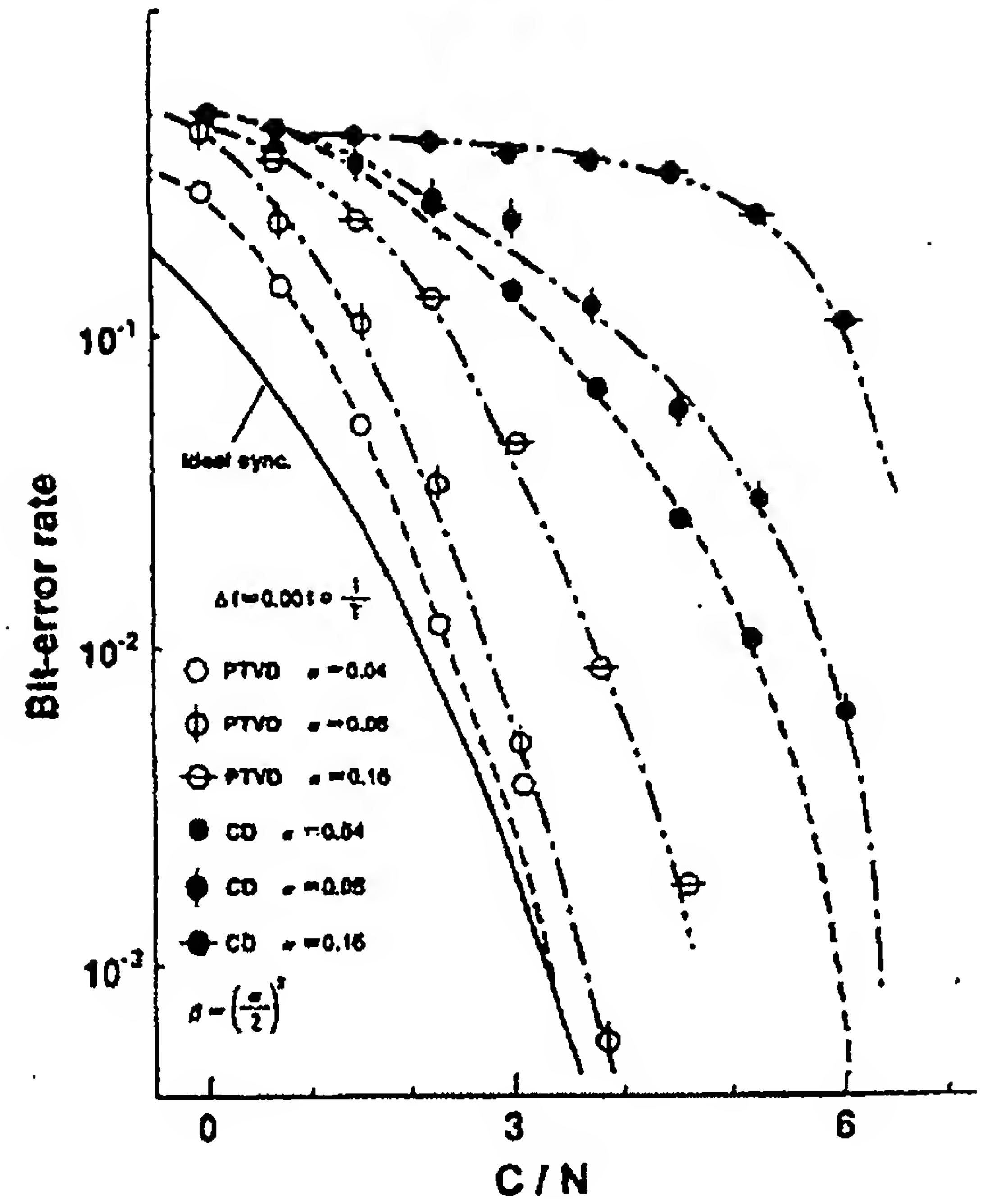

Fig. 6. Bit-error rate for phase-tracking Viterbi demodulator.

It is assumed that the known signal is sufficiently long and the phase uncertainty can be completely eliminated.

Consider the situations where the number of error bits in a slot exceeds 200 and 50 after the convolution decoding. They are called the 20-percent error slot and the 5-percent error slot, respectively. In the very low SNR condition, where the desynchronization and the cycle slip occur frequently and continuously, it is difficult to distinguish between the desynchronization and the cycle slip. Then it is practically impossible to measure the number of times for the cycle slip. Consequently, the generations of the 20-percent and 5-percent error slots are used to represent the generation of the cycle slip.

The generation of the 20-percent error slot indicates that a cycle slip or a desynchronization is produced soon after receiving the known signal. The generation of the 5-percent error slot indicates that a cycle slip is produced at a certain elapse of time after receiving the known signal. In other words, when a large number of 20-percent error slots are generated, it cannot be expected that the overall error rate is greatly improved even if the slot length is decreased. When a large number of 5-percent error slots are generated but there are fewer than 20-percent error slots, it can be expected that the overall error rate is improved by carefully determining the slot length.

5.2. Result of simulation (error rate performance)

Figures 5 and 6 show the error rate performances of the proposed Viterbi demodulator with the phase synchronization function (PTVD) and the conventional demodulator (CD) composed of a serial connection of the coherent detector and the maximum-likelihood sequence estimator, respectively.

In the measurement of the error rate performance, it is assumed that the constant frequency offset is produced in each slot (immediately after receiving the known signal). In other words, the phase varies in a ramp form. This corresponds to the situation in the satellite communication and deep-space exploration where the satellite or the explorer modifies the posture or orbit while executing transmission/reception. It also corresponds to the situation in mobile satellite communication

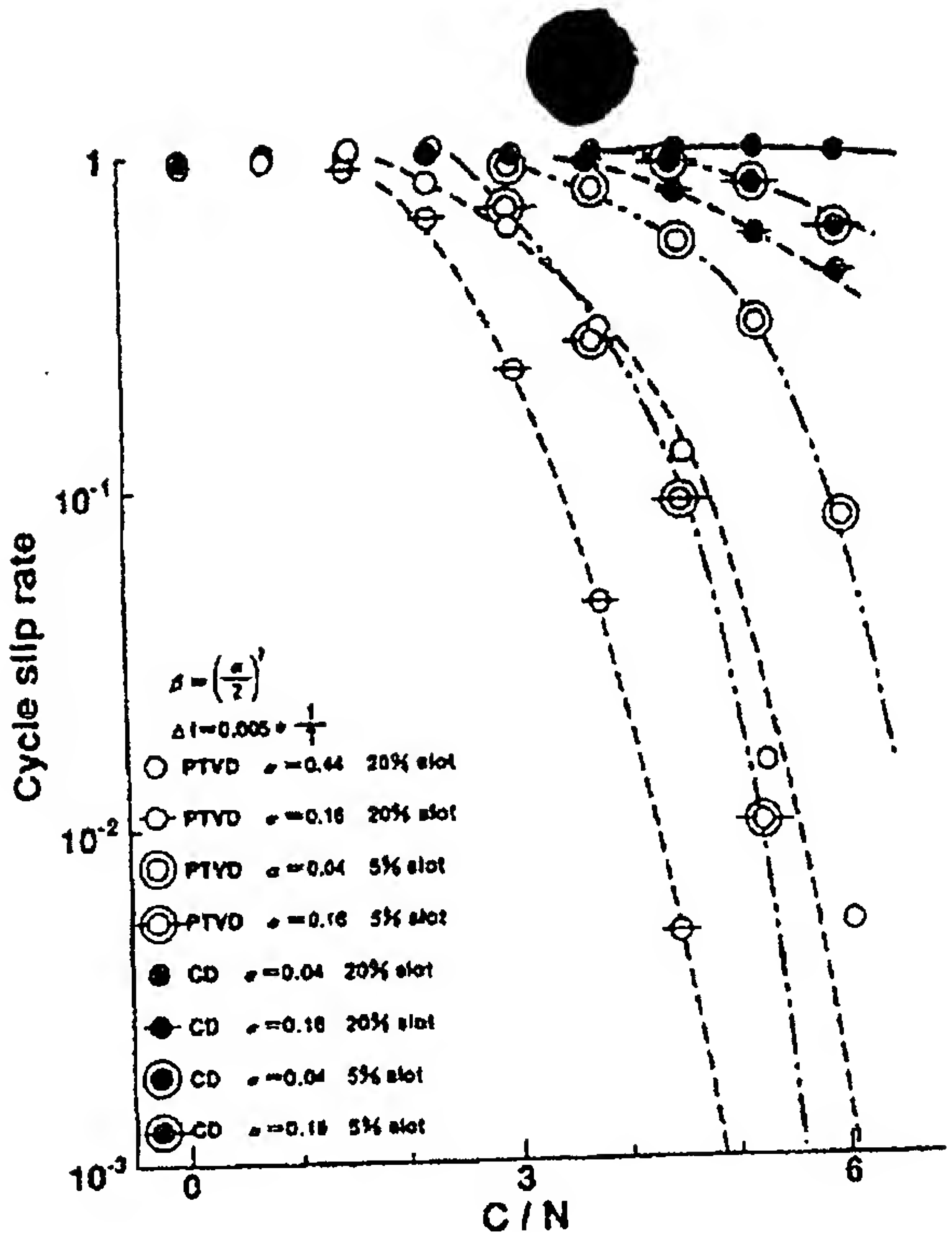

Fig. 7. Cycle slip for phase-tracking Viterbi demodulator.

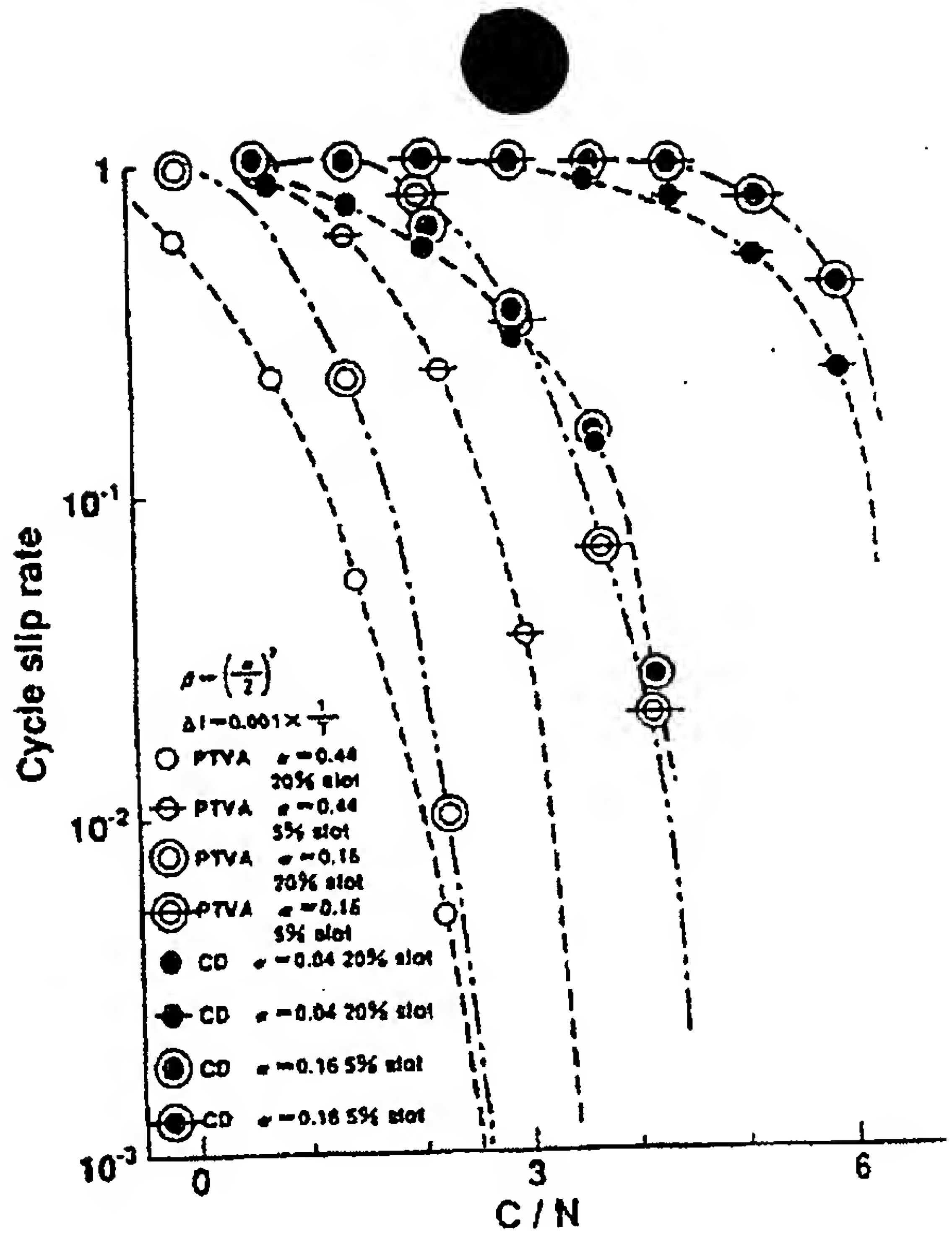

Fig. 8. Cycle slip for phase-tracking Viterbi demodulator.

in which the mobile station changes the direction of motion. In such cases, the receiver must follow the Doppler-shift and other changes produced by the variation of the relative velocity, with a sufficient tolerance. The figures show the performance for various frequency offsets. The frequency offsets in Figs. 5 and 6 correspond to 0.5 percent and 0.1 percent of the transmission rate, respectively.

α and β, which are the coefficients of the loop filter, are set as the same values in the two systems. The relation $(\alpha/2)^2 = \beta$ is maintained between α and β. This relation is the condition for the critical damping in the secondary tracking detector loop shown in Fig. 2(a).

It is seen from the figure that the conventional system can hardly maintain the synchronization when CNR goes below 3 dB, resulting in the error rate of 0.5 while the proposed system maintains a sufficiently low error rate. In other words, it is seen from the figures that the proposed system exhibits a much lower error rate than the conventional system, especially when SNR is low. In Fig. 5, the proposed system exhibits the best performance for $\alpha = 0.08$. In other words, the optimal performance for the frequency variation of some 0.5 percent of the transmission rate, from the viewpoints of handling the phase variation and the noise immunity, is given by this parameter. This performance cannot be achieved in the conventional system; however, the coefficient of the loop filter may be adjusted. When the frequency offset is 0.1 percent of the transmission rate, on the other hand, the best performance is obtained in either system for $\alpha = 0.04$. Especially, in the proposed system, the overall error rate which is almost the same as that of the ideal synchronization performance (hard-wired carrier) is obtained.

Figures 7 and 8 show the generation probabilities of the 20-percent and the 5-percent error slots, respectively, as functions of CNR. It is again seen from the figures that the proposed system has a better noise immunity and handles the frequency offset better. Especially, it is seen that the generation probabilities for the 20-percent error slot and the 5-percent error slot differ greatly. This indicates that the overall error rate will be further improved by reducing the slot length or other elaborations. In all cases of Figs. 5 to 8, the error rate and the probability of cycle slip rapidly decrease when

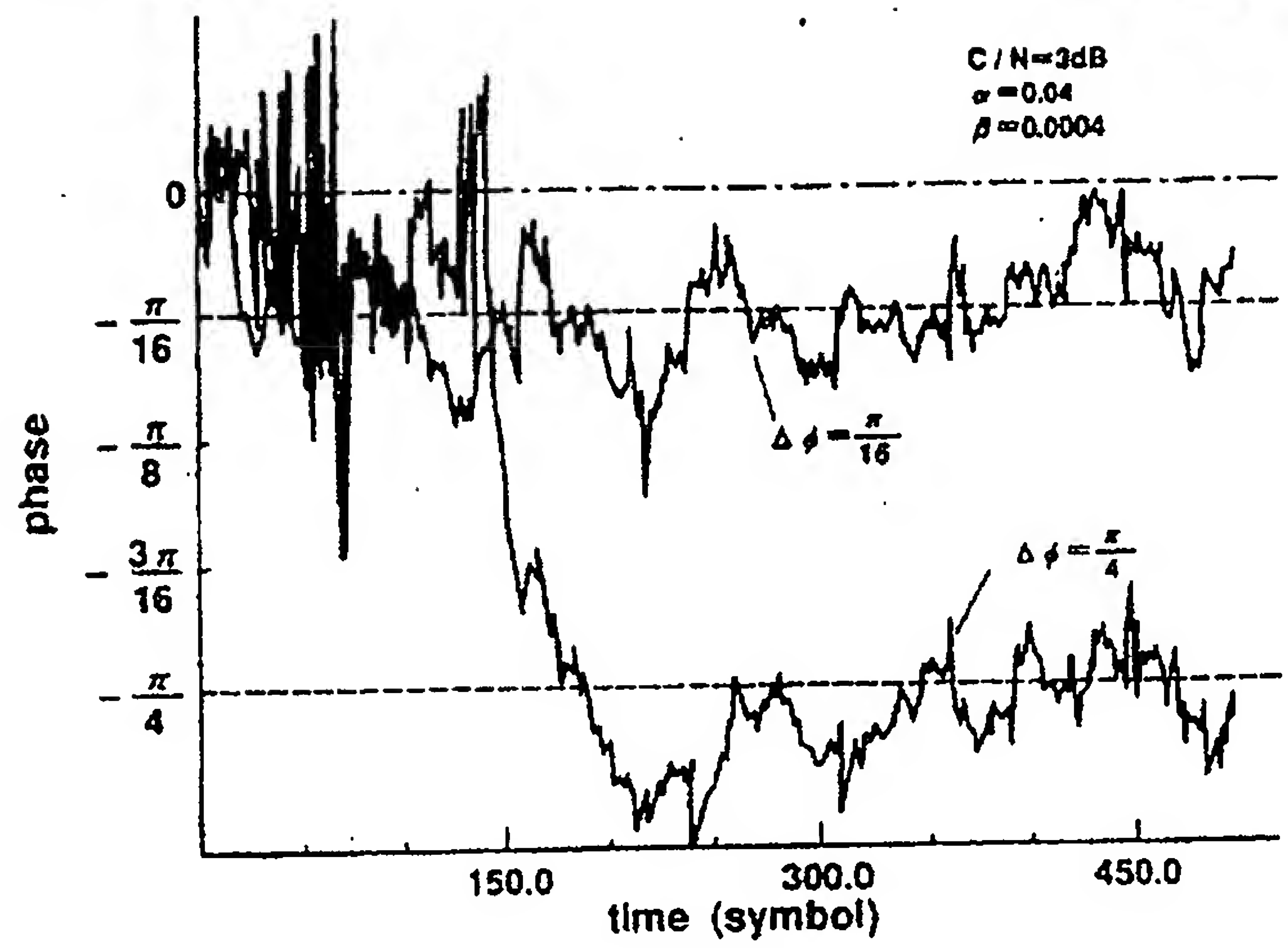

Fig. 9. Phase-step response.

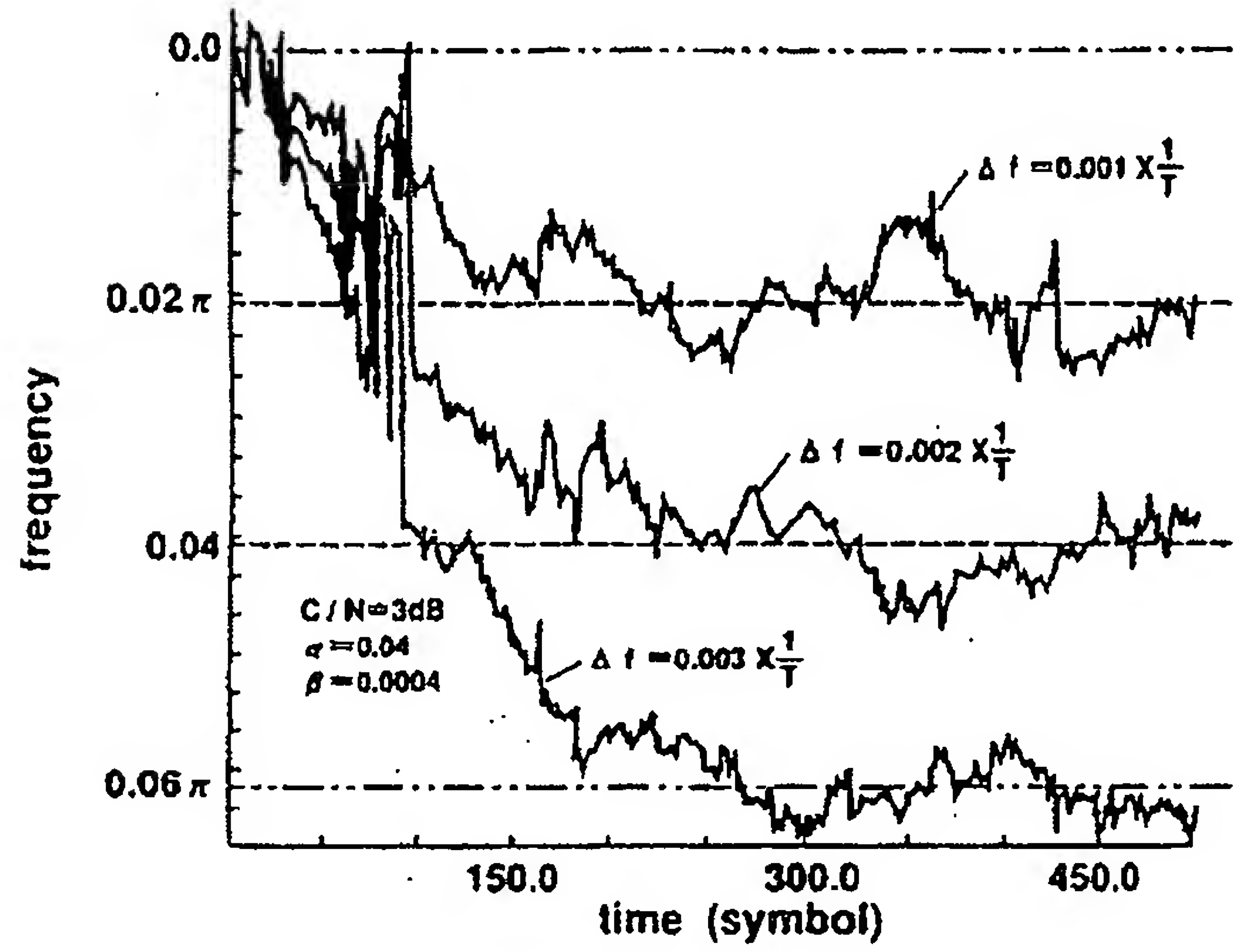

Fig. 10. Frequency step response.

the CNR reaches 6 to 8 dB. In this simulation, the transmission is tried for more than 1 million symbols in either case. For the high SNR (>8 dB), no error or cycle slip was found. In other words, by applying the proposed method, the SNR, for which the receiver is to operate, can be set much lower.

5.3. Result of simulation (synchronization performance)

In order to verify that the satisfactory pull-in of the synchronization can be realized for the low SNR, the initial synchronization performance of the proposed

system is examined. Figures 9 and 10 show the phase synchronization process in the proposed system. Figure 9 is the case where the stepwise phase variation is imposed, and Fig. 10 is the case where the stepwise frequency variation is imposed. Figure 9 shows which memorized state has the highest likelihood, among the memory phases for the states as a function of the time. Figure 10 shows the value memorized in the state corresponding to the maximum likelihood, among the loop-filter registers for the states (frequency memory).

When the state is selected incorrectly, there is a possibility that the phase and the frequency offset may be outputted, which differs from those used in the demodulation of the actual output demodulated signal. This can be considered as the case where a large phase or frequency variation occurs, immediately after the stepwise variation of the phase or frequency. It is seen from the figures that the control progresses toward the goal value, and the synchronization is gradually established.

One hundred trials are made for the forementioned synchronization pull-in process, and the same pull-in performance is obtained for all trials. Thus, it is verified from those figures that a satisfactory phase pull-in performance is obtained even for the low SN condition.

6. Conclusions

This paper proposed and examined the Viterbi demodulation system with the phase-tracking function for the phase modulated convolution code. It is shown that the proposed system operates well for the low SN and a large carrier phase variation, where it has been difficult or practically impossible in the conventional system to maintain the phase synchronization. Especially, it is shown that the proposed system is better than the conventional system in the very adverse condition with the CNR below 3 dB. The basic performances of the proposed system are demonstrated.

The following problems are left for the future study. The pull-in process of the system proposed in this paper should be analyzed in more detail. The performance, when the proposed method is applied to various modulation/coding system such as CPM and TCM should be evaluated and discussed. The method should be compared to the case where the error-correcting code is applied to the coherent detection by an approach different from the method in this paper [18].

REFERENCES

1. V. K. Bhargava et al. Digital Communications by Satellite. John Wiley & Sons (1981).
2. H. Suzuki et al. MODEM and FEC LSIs for Highly Functional Compact Earth Station. IEEE Proc. GLOBECOM '87, 8-3 (1987).
3. K. Murota et al. GMSK modulation for digital mobile telephony. IEEE Trans. Commun., COM-29, 7, pp. 1044-1050 (1981).
4. T. Aulin et al. Continuous phase modulation—Part 1: Full response signaling. IEEE Trans. Commun., COM-29, 3, pp. 196-209 (1981).
5. T. Aulin et al. Continuous phase modulation—Part 2: Partial response signaling. Ibid., pp. 210-225.
6. C.-E. Sundberg. Continuous-phase modulation. IEEE Commun. Mag., 24, 4, pp. 25-38 (April 1986).
7. J.-E. Stejernvall et al. Performance of a Cellular TDMA System in Severe Time Dispersion. IEEE Proc. GLOBECOM '87, 21-7.
8. G. Ungerboeck. Trellis-coded modulation with redundant signal sets—Part I and Part II. IEEE Commun. Mag., 25, 2, pp. 5-21 (Feb. 1987).
9. CCITT Draft Recommendation V. 33.
10. G. D. Forney, Jr. Maximum likelihood sequence estimation of digital sequences in the presence of intersymbol interference. IEEE Trans. Inf. Theory, IT-18, 3, pp. 363-378 (May 1972).
11. A. Baier et al. Bit Synchronization and Timing Sensitivity in Adaptive Viterbi Equalizer for Narrow-band TDMA Digital Mobile Systems. IEEE Proc. VTC '88, pp. 377-384.
12. G. Ascheid et al. Performance Analysis of a Decision-Directed Feedback Carrier Synchronization System for Bandwidth-Efficient Modulation. IEEE Proc. GLOBECOM '84, 8-8.
13. B. A. Mazur et al. Demodulation and carrier-synchronization of multi-h phase codes. IEEE Trans. Commun., COM-29, 3, pp. 257-266 (March 1981).
14. G. Ungerboeck. New Application for the Viterbi Algorithm: Carrier-Phase Tracking in Synchronous Data-Transmission Systems. IEEE Proc. NTC '74, pp. 734-738.
15. J. B. Thorpe et al. A hybrid phase/data Viterbi demodulator for encoded CPFSK modulation. IEEE Trans. Commun., COM-33, 6, pp. 535-542 (June 1985).
16. M. Serizawa et al. Phase-tracking Viterbi demodulator. Electron. Lett., 25, 12, pp. 792-793 (June 1989).

17. M. Serizawa et al. Phase-Tracking Viterbi Demodulator. 1989 IEICE Spring Meeting, B-85.

18. Y. Matsumoto et al. A New Carrier Recovery Circuit for Mobile Satellite Communications. IEEE Proc. VTC'92, pp. 866-869.

AUTHORS (from left to right)



Mutsumu Serizawa graduated in 1981 with a B.E. degree from the Department of Electrical Engineering, Keio University, and affiliated with Toshiba Corporation. In the Research and Development Center, he is engaged in research and development in mobile communication equipment and systems as well as satellite communication equipment. He was a Technical Visitor, Thomson Plant Co., Angier Laboratory, for one year (1982), and at Rutgers State University, WINLAB (Wireless Information Network Laboratory), NJ, U.S.A., for half a year from 1992. At present, he is a Research Chief, Toshiba Research and Development Center, Information Communication Systems Laboratories, and a member of the IEEE (U.S.A.).

Shigeru Asakawa graduated with a B.S. and a Ph.D. degree in 1962 and 1982, respectively, from Osaka University and affiliated with Toshiba Corporation. In the Research and Development Center, he is engaged in research and development mostly on digital communication systems such as speech/image PCM codec, multiple access in satellite communication, digital radio-call system, and digital modulation/demodulation systems. He was with Toshiba London Office from 1982 to 1985 and was engaged in technical negotiations with European countries. Director of Electronic Equipment Laboratory, Toshiba Research and Development Center in 1987 and Director of Toshiba Kansai Research Laboratory in 1991. At present, he is with Kansai New Technical Research laboratory. He is a Senior Member of the IEEE (U.S.A.) and the IEE (U.K.).

Junzo Murakami graduated with a B.E. degree in 1961 from the Department of Electrical Engineering, Faculty of Engineering, University of Tokyo, and obtained his Ph.D. later. He joined Toshiba Corporation. In the Research and Development Center he engaged in research and development of low-noise microwave receiver, radio communication system, TV ghost canceller, and digital modulation system. At present, he is a Fellow at Toshiba Research Consultation Co. He received Paper Awards in 1961 and 1984, I.E.I.C.E.J. He is a member of the IEEE (U.S.A.).